Article

Gene Expression of *Haloferax volcanii* on Intermediate and Abundant Sources of Fixed Nitrogen

Sungmin Hwang [1], Nikita E. Chavarria [2], Rylee K. Hackley [1,3], Amy K. Schmid [1,3,4,*] and Julie A. Maupin-Furlow [2,5,*]

[1] Department of Biology, Duke University, Durham, NC 27708, USA; sungmin.hwang@duke.edu (S.H.); rylee.hackley@duke.edu (R.K.H.)
[2] Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL 32611, USA; nikita.chavarria@utrgv.edu
[3] University Program in Genetics and Genomics, Duke University, Durham, NC 27708, USA
[4] Center for Genomics and Computational Biology, Duke University, Durham, NC 27708, USA
[5] Genetics Institute, University of Florida, Gainesville, FL 32611, USA
* Correspondence: amy.schmid@duke.edu (A.K.S.); jmaupin@ufl.edu (J.A.M.-F.)

check for updates

Abstract: *Haloferax volcanii*, a well-developed model archaeon for genomic, transcriptomic, and proteomic analyses, can grow on a defined medium of abundant and intermediate levels of fixed nitrogen. Here we report a global profiling of gene expression of *H. volcanii* grown on ammonium as an abundant source of fixed nitrogen compared to ʟ-alanine, the latter of which exemplifies an intermediate source of nitrogen that can be obtained from dead cells in natural habitats. By comparing the two growth conditions, 30 genes were found to be differentially expressed, including 16 genes associated with amino acid metabolism and transport. The gene expression profiles contributed to mapping ammonium and ʟ-alanine usage with respect to transporters and metabolic pathways. In addition, conserved DNA motifs were identified in the putative promoter regions and transcription factors were found to be in synteny with the differentially expressed genes, leading us to propose regulons of transcriptionally co-regulated operons. This study provides insight to how *H. volcanii* responds to and utilizes intermediate vs. abundant sources of fixed nitrogen for growth, with implications for conserved functions in related halophilic archaea.

Keywords: halophilic archaea; nitrogen metabolism; ammonium uptake; transcriptome; alanine transport; transcription factor; Lrp-type

1. Introduction

Transcription is a primary mechanism that enables organisms to regulate gene expression to produce cellular factors in response to changing environmental conditions. Traditional methodologies that detect differential gene expression, like Northern blots and quantitative reverse transcription-polymerase chain reaction (PCR) [1,2], are useful to characterize specific target genes. In concert, global transcriptional analyses using expression microarray, and most recently RNA-seq, are widely used to analyze the gene expression level of bacteria and eukaryotes in response to various stimuli and environmental conditions [3–6]. Transcriptomic analysis is also used as a pivotal tool to identify how archaeal gene expression is regulated under environmental stress or nutrient fluctuations [7–10].

Nitrogen, one of the most common elements on the earth, is present in different redox states from nitrate (NO_3^-) to ammonium (NH_4^+). The majority of nitrogen in biomolecules is present in reduced forms. Biologically, nitrogen comprises amide groups in amino acids and functional groups in nucleic

acids, which are essential for growth and reproduction [11]. Interconversion of the different forms of nitrogen, termed N-cycle, is coordinated by biogeochemical redox processes. Nitrification, the oxidative conversion of ammonium into nitrite (NO_2^-) and further nitrate (NO_3^-), is mediated by a variety of bacteria and archaea in soils and oceans [12,13]. On the other hand, denitrifying microorganisms reduce NO_3^- to nitrogenous compounds, such as NO_2^-, NO, N_2O and N_2 through denitrification processes [14]. N_2 can be converted to ammonium by diazotrophs and archaeal methanogens through the assimilatory mechanism of nitrogen fixation; while ammonium is also generated from NO_3^- and NO_2^- by dissimilatory reduction through anaerobic respiration and by assimilatory reduction for cell carbon biosynthesis [15–17]. Considering that N-cycle compounds impact natural environments (e.g., N_2O is a potent greenhouse gas) and human health (e.g., nitrate consumption is relevant to cancer and adverse reproductive outcomes), better understanding of nitrogen metabolism across diverse microbial groups will inform on the role of different environments in the global N-cycle [18,19].

Given the tremendous interest in understanding how archaea respond to and control nitrogen cycles, their transcriptional and post-transcriptional responses to nitrogen availability have been probed at the gene-specific and global levels. *Sulfolobus acidocaldarius*, a member of Crenarchaeota, encodes a leucine-responsive regulatory protein-like family transcriptional regulator Sa-Lrp [20]. In vitro Sa-Lrp binds specifically to the promoters of genes involved in nitrogen metabolism, such as glutamine synthetase (*glnA*-1, *glnA*-2, and *glnA*-3) and glutamate synthase (*gltB*) [20], suggesting that Sa-Lrp may regulate transcription of the GS/GOGAT pathway depending on N-signals. Transcriptional responses to N-conditions are best understood in the Euryarchaeota: the nitrogen-fixing methanogens [21,22] and the denitrifying halophile *Haloferax mediterranei* [8,23,24]. The methanogens use a master transcriptional repressor NrpR, secondary transcriptional activator NrpA, and small RNAs (sRNAs) to control the expression of glutamine synthetase (*glnA*) and nitrogenase (*nif*) genes in response to N-sources [22]. *H. mediterranei* mediates transcriptome level responses that are consistent with its capacity for denitrification. For instance, genes involved in the assimilatory reduction of nitrate to nitrite (*nasA*) and nitrite to ammonium (*nasD*) are up-regulated in the presence of nitrate [25,26]. In addition, the transcript levels of the high affinity Amt-type ammonia transporter and PII (GlnK) regulators are up during N-starvation in *H. mediterranei* [27]. The regulatory network responsible for controlling the N-responses in non-methanogenic haloarchaea, however, is poorly understood; sRNAs are implicated in control of the Amt/PII and glutamate dehydrogenase (*gdhA1*) genes [24], but a master NrpR or NrpA-type transcriptional regulator is not conserved.

Here, we use global transcriptome profiling to examine the response of *Haloferax volcanii* to L-alanine vs. ammonium as a nitrogen (N-) source. *H. volcanii*, a close relative to *H. mediterranei*, is a model archaeon originally isolated from the Dead Sea (salinity ca. 340 g/L) [28], where N-sources are diverse and ammonia concentrations have gradually increased over time due to pollution [29]. L-alanine was chosen as the intermediate N-source based on prior knowledge that: i) *H. volcanii* can use L-alanine as an N-source [30]; and ii) the methanogenic archaeon *Methanococcus maripaludis* responds at the transcriptional level to L-alanine as an intermediate N-source [31].

Our findings provide insight into how halophilic archaea, such as *H. volcanii*, respond to intermediate (L-alanine) vs. abundant (ammonium) N-sources at the global level. Using microarray analysis, gene expression profiles were detected that enabled us to map putative transporters and metabolic pathways specific to the two N-sources. In synteny with these N-regulated operons, a putative *cis*-regulatory binding motif and an Lrp-like transcription factor were identified as candidates for controlling a transcriptional response to intermediate vs. abundant N-sources.

2. Results and Discussion

2.1. Genome-Wide Expression Analysis under Different Nitrogen Sources

H. volcanii H26 strain was analyzed for growth on glycerol minimal medium (GMM) with the N-source of 10 mM ammonium chloride (simplified to 'ammonium or NH_4^+') or 10 mM L-alanine

(Figure A1). While the doubling time of H26 was similar under each condition (7.2 ± 0.14 h with L-alanine and 8.2 ± 0.19 h with ammonium; 1.1-fold difference), the final cell density in stationary phase (i.e., carrying capacity) increased 1.7-fold ($p < 2.68 \times 10^{-15}$, unpaired, two-sided *t*-test) when H26 was grown on L-alanine compared to ammonium (Figure A1). This distinction in cell density suggests that L-alanine may serve as a carbon source, in addition to serving as a N-source. The results also suggest that the cells may differ in their transcriptional response to these two N-sources.

To determine how cells respond to N-source shifts at the level of global gene expression, total RNA of H26 was isolated from log-phase cells grown with L-alanine or ammonium and the transcripts were analyzed by microarray. Expression was reproducible across biological replicates within each condition (correlation R^2 = 0.95–0.97 and 0.90–0.97 for ammonium and L-alanine, respectively, see also GitHub repository, URL given in Methods). In addition, average expression across all genes across the two conditions was R^2 = 0.90, suggesting good reproducibility of growth conditions with limited, condition-specific transcriptome changes. Differential expression was calculated for all genes by statistical analysis of the detected hybridization signals (Figure 1A, Methods). Thirty genes were identified as significantly differentially expressed between the two conditions, and those genes were clearly divided into two groups (Figure 1B): 10 genes were down-regulated and the remaining 20 genes were up-regulated in GMM with L-alanine versus ammonium (Table 1). These 30 differentially expressed genes were enriched for functions coding for amino acid metabolism and transport according to the eggNOG database for orthology mapping and functional classification [32] (*p* value ≤0.05, hypergeometric test). This functional enrichment of differentially expressed genes is in line with the biological role of nitrogen. RT-qPCR was performed to validate the genes identified by the microarray (Figure 2A). All the genes measured by RT-qPCR were up- or down-regulated on L-alanine/NH$_4$Cl growth condition in concordance with the microarray data. Across all 6 genes measured, the two methods were strongly correlated (Spearman's ρ = 0.71), with some gene-specific variation (Figure 2B). Taken together, these independent experimental results suggest that the genes identified by microarray are differentially expressed depending on the N-source.
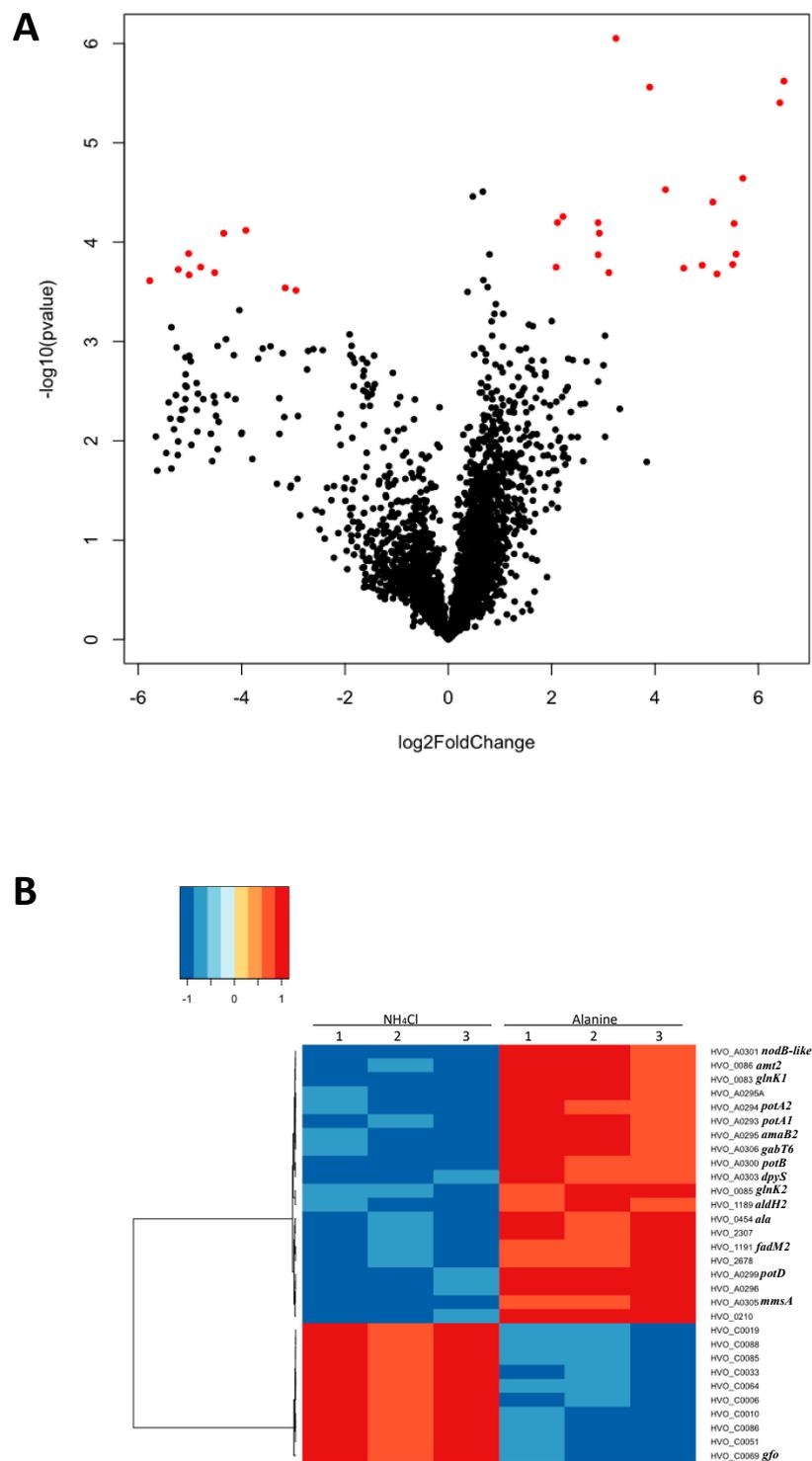
Figure 1. Microarray data analysis. (**A**) Each gene analyzed is visualized as a dot in the volcano plot. Red dots indicate genes whose log2-transformed expression ratio met the 2-fold and the false discovery rate (FDR) cutoffs. (**B**) The heat map represents the results of hierarchical clustering of genes differentially expressed in the growth conditions supplemented ʟ-alanine compared to ammonium. Each gene in each row is labeled at the right side of the heat map with common gene names. Column numbers indicate the biological replicates under each condition. The dendrogram at the left shows the groups of co-expressed genes that resulted from clustering analysis. The color key at the top left indicates the correlation matrix (from −1 to 1 scale) of the log2 gene expression ratios in ʟ-alanine: ammonium. See analysis details in Materials and Methods.

Table 1. Genes differential expression on L-alanine compared to ammonium in minimal medium.

Uniprot Annotation [1]; Proposed Function [2]	Gene	Log2 Fold Change	SD [3]	SD Error
ABC-type transport system periplasmic substrate-binding protein (probable substrate spermidine/putrescine); L-alanine transport	HVO_A0299, *potD* [5]	6.495	0.177	0.102
Polysaccharide deacetylase family protein	HVO_A0301, *nodB*-like	6.416	0.280	0.161
(Methyl)malonate-semialdehyde dehydrogenase [EC:1.2.1.27]	HVO_A0305, *mmsA* [5]	5.701	0.310	0.179
Luciferase family protein; oxidoreductase	HVO_A0295A	5.569	0.396	0.228
ABC-type transport system permease protein (probable substrate spermidine/putrescine); L-alanine transport	HVO_A0300, *potB* [4]	5.530	0.609	0.351
Amidase (hydantoinase/carbamoylase family); β-ureidopropionase / *N*-carbamoyl-L-amino-acid hydrolase [EC:3.5.1.6/3.5.1.87]	HVO_A0295, *amaB2* [4]	5.504	0.637	0.368
Putative allantoinase; dihydropyrimidinase [EC 3.5.2.2]	HVO_A0303, *dpyS* [5]	5.200	0.771	0.445
ABC-type transport system ATP-binding protein (probable substrate spermidine/putrescine); L-alanine transport	HVO_A0294, *potA2* [4]	5.119	0.293	0.169
ABC-type transport system ATP-binding protein (probable substrate spermidine/putrescine); L-alanine transport	HVO_A0293, *potA1* [4,5]	4.913	0.656	0.379
Pyridoxal phosphate-dependent aminotransferase	HVO_A0306, *gabT6* [4,5]	4.555	0.508	0.293
Nitrogen regulatory protein P-II	HVO_0085, *glnK2* [4]	4.203	0.241	0.139
Transport protein (Probable substrate ammonium)	HVO_0086, *amt2*	3.898	0.165	0.095
Short-chain family oxidoreductase	HVO_A0296	3.244	0.106	0.061
Alanine dehydrogenase [EC 1.4.1.1]	HVO_0454, *ala* [4]	3.107	0.537	0.31
Nitrogen regulatory protein P-II	HVO_0083, *glnK1* [4,5]	2.922	0.303	0.175
Proline dehydrogenase [EC:1.5.5.2]	HVO_1191, *fadM2* [4]	2.901	0.280	0.161
Uncharacterized protein (arCOG09242)	HVO_0210	2.898	0.335	0.193
HAD superfamily hydrolase with phosphoglycolate phosphatase-like, domain 2 (IPR023198)	HVO_2307	2.22	0.270	0.156
Uncharacterized protein of arCOG06214	HVO_2678	2.113	0.231	0.133
Aldehyde dehydrogenase; delta-1-pyrroline-5-carboxylate dehydrogenase [EC:1.2.1.3]	HVO_1189, *aldH2*	2.086	0.243	0.140
Glucose-fructose oxidoreductase	HVO_C0069, *gfo*	−2.949	0.167	0.096

Table 1. *Cont.*

Uniprot Annotation [1]; Proposed Function [2]	Gene	Log2 Fold Change	SD [3]	SD Error
HTH domain protein	HVO_C0051 [5]	−3.156	0.414	0.239
Uncharacterized protein	HVO_C0010	−3.918	0.376	0.217
Uncharacterized protein	HVO_C0033	−4.347	0.541	0.313
Small CPxCG-related zinc finger protein	HVO_C0086	−4.522	0.381	0.220
UPF0395 family protein	HVO_C0006	−4.792	0.715	0.413
Homolog to virus protein eHPD7_00180	HVO_C0085	−5.018	0.585	0.338
ISH4-type transposase ISHvo5	HVO_C0064	−5.027	0.662	0.382
hypothetical protein	HVO_C0019	−5.227	0.778	0.449
hypothetical protein	HVO_C0088	−5.779	0.837	0.483

[1] Uniprot annotation as of May 14, 2019.; [2] Proposed by this study; [3] SD, standard deviation of the log2FoldChange value; [4] Genes of the enriched eggNOG category of amino acid metabolism; [5] Promoter contains predicted DNA binding motif.
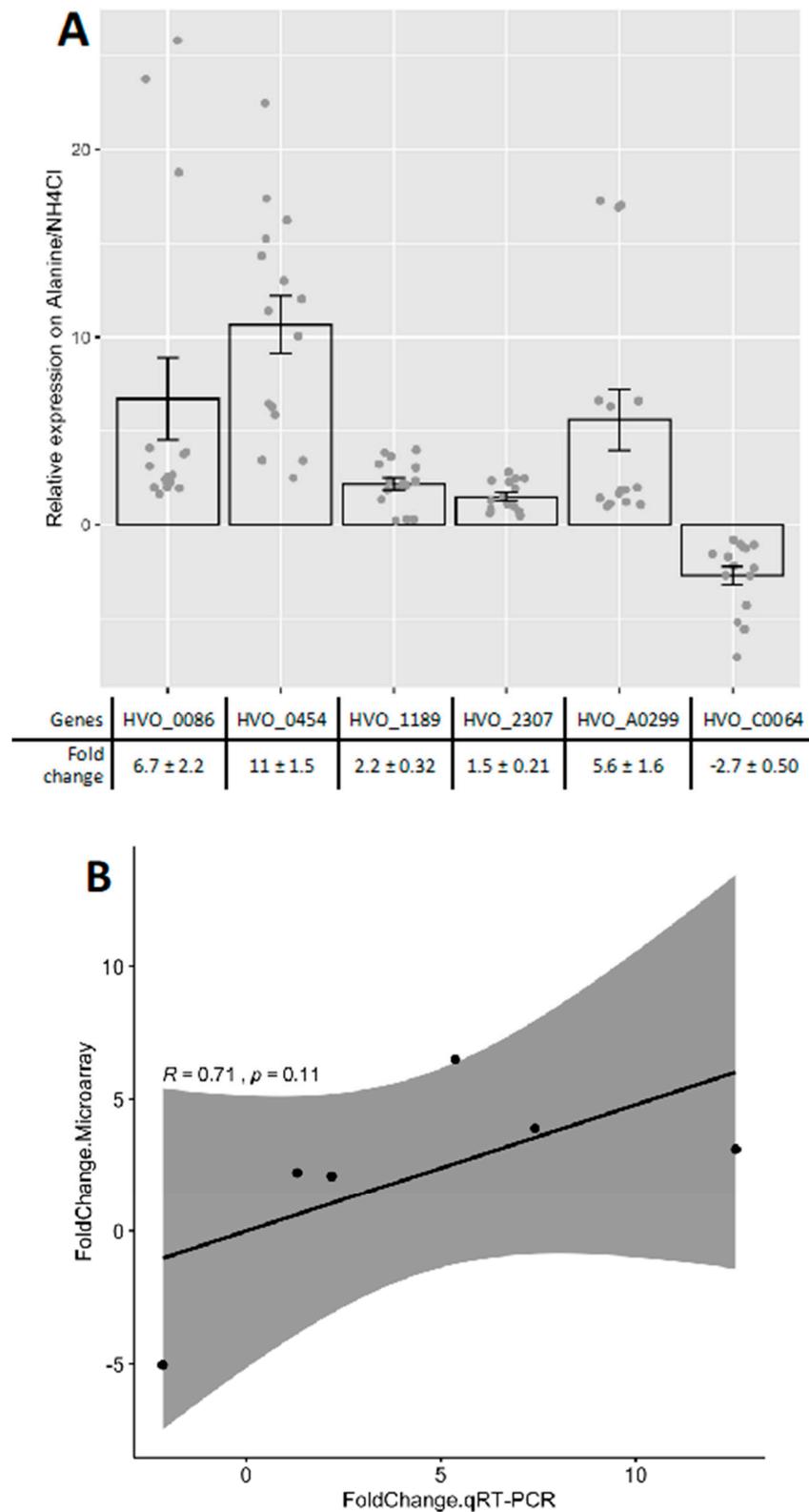
Figure 2. Validation of gene expression profiling by reverse transcriptional quantitative PCR (RT-qPCR). (**A**) Representative genes are from Table 1 and total *n* = 12 (biological quadruplicate and technical triplicate) was applied for the RT-qPCR. Error bars, standard error of the mean. (**B**) The expression magnitude of each genes between microarray and RT-qPCR was averaged and compared by Spearman's correlation method.

Notably, 8 of the 10 significantly down-regulated genes are predicted to encode hypothetical proteins with unknown function. The two exceptions are ISH4-type transposase and glucose-fructose oxidoreductase gene homologs (Table 1). All 10 genes down-regulated are encoded on plasmid pHV1. Surprisingly, the mean log2-fold value (alanine vs. ammonium) for all 75 genes encoded on pHV1 was −4.1 in this microarray analysis regardless of significance cutoff. In contrast, the mean expression for all genes on other genomic elements were: main chromosome 0.23; pHV2, 0.21; pHV3, −0.18; pHV4, 0.05. All the genes on pHV1 were down-regulated, whereas up- (66%) and down-regulated genes (34%) were evenly represented on the other chromosomal elements (Table S1). Gene expression level of one of the biological replicates in ʟ-alanine (H26_ala_C in Table S1) differed from the other two with a *p* value $< 2 \times 10^{-16}$ (Pairwise Wilcoxon Rank Sum Test with Bonferroni Correction). However, the average log2-fold value (alanine vs. ammonium) of all pHV1 genes was still less than zero without this single biological replicate included in the average across biological replicates (H26_ala_C). Down-regulation of pHV1 genes was therefore not driven by differences between pHV1 expression across the biological replicates. One alternative explanation is that pHV1 gene expression was highly dependent on the replicon copy number and that N-source impacted this copy number through plasmid loss, rearrangement or altered polyploidy, as have been previously observed in *H. volcanii* and other halophiles [33–35]. To determine the relative copy number of the main chromosome and plasmid pHV1 in the two different growth conditions, quantitative PCR (qPCR) was performed with genomic DNA. The normalized amplicon level of two pHV1 genes, HVO_C0033 and HVO_C0069, was found to be similar to that of an internal control *rpl16* (HVO_0484) located on the main chromosome (Figure A2, Table S1). A previous study reported that the copy number of the main chromosome is 1.6-fold higher than that of pHV1 in exponential phase in *H. volcanii* [36]. The copy number discrepancy between the current and previous study might be caused by different growth conditions: complex vs. glycerol minimal medium. Further analysis of transcript levels by RT-qPCR showed HVO_C0033 and HVO_C0069 to be reduced in expression when cells were grown in ʟ-alanine relative to ammonium (Figure A2B), in agreement with the microarray data. Taken together, these results suggest that the down-regulated gene expression in pHV1 during exposure to ʟ-alanine vs. ammonium is not due to altered plasmid copy number. However, we cannot rule out that pHV1 was randomly lost during growth of cells for the microarray experiment but was maintained in cultures grown for the qPCR experiments (Figure A2B). The *H. volcanii* genome is highly plastic [35], and random gene loss on megaplasmids during microarray experiments has been observed in other halophiles with plastic genomes [33]. The mechanism involved in coordinated down-regulation across the entire pHV1 plasmid therefore remains unclear.

2.2. Putative Amino Acid Transport Systems were Upregulated on ʟ-Alanine

Of the 20 genes that were found upregulated at least 2-fold on ʟ-alanine compared to ammonium as the N-source (Figure 3), seven were associated with transport (Figure 4). Included in this list were the Amt-type high affinity ammonium transporter (Amt2) and PII regulator (GlnK1 and GlnK2 or GlnK1/2) genes (Figure 3, Table 1). The PIIs are predicted to regulate transport of ammonium by Amt2 based on analogy to *E. coli* [37]. *H. mediterranei* undergoes a similar increase in *amt-glnK* transcript abundance when starved for nitrogen (N) [27]. Thus, in halophilic archaea, the Amt-transport system is upregulated by shifts to intermediate sources of fixed nitrogen in addition to N-limitation presumably to scavenge ammonium from the environment. The other transport system upregulated on ʟ-alanine was the ABC-type transporter system PotA1, PotA2, PotB and PotD (PotA1A2BD) (Figure 4). While annotated as a potential spermidine/putrescine transporter (UniProt, April 10, 2019 update), the PotB permease of this system had a MetI-like transmembrane domain (IPR000515) related to the D-methionine ABC transporter of *E. coli* [38]. The high expression of *potA1A2BD* on ʟ-alanine (*potB* and *potD* were among the most highly upregulated transcripts, Table 1), combined with the relationship of PotB to MetI, suggest a function of the encoded ABC-type system in the transport of amino acids such as ʟ-alanine.
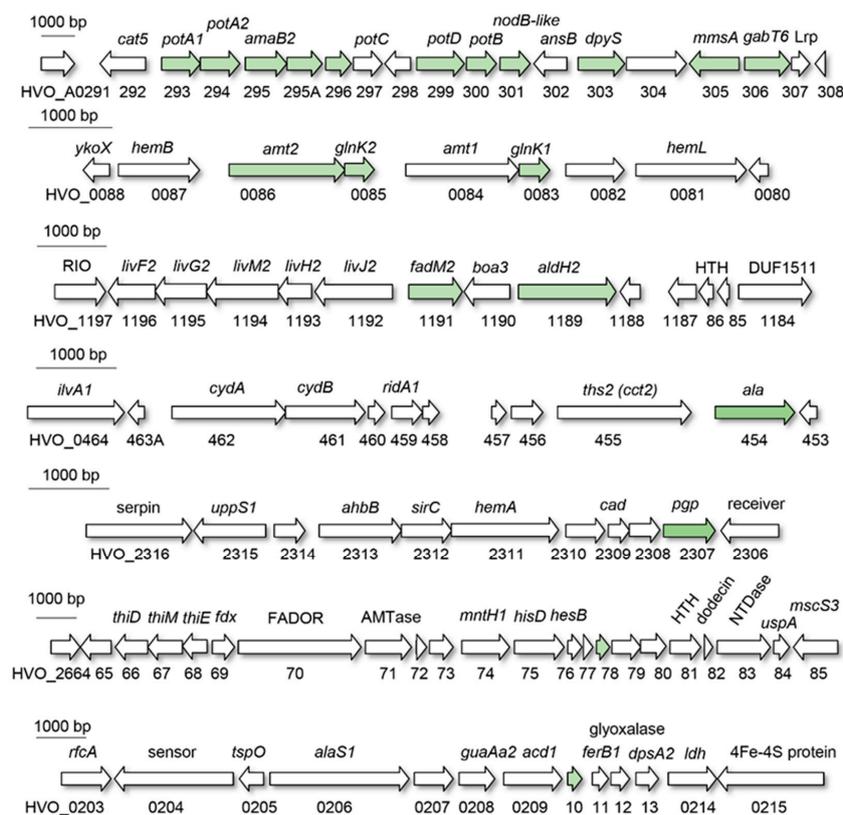
Figure 3. Operon organization of *H. volcanii* chromosomal and plasmid pHV4-based genes that are highly expressed on ʟ-alanine compared to ammonium as an N-source. Arrows indicate open reading frames (ORFs) with those in green indicating genes that have transcripts more abundant on ʟ-alanine compared to ammonium. Annotated gene names and locus tag numbers are indicated above and below the ORF, respectively.

2.3. Upregulation of Metabolic Systems

Seven of the 20 genes upregulated on ʟ-alanine compared to ammonium were mapped to metabolic pathways. Of these upregulated genes, HVO_0454 (*ala*) shares 46% identity and 60% similarity in amino acid sequence to the ʟ-alanine dehydrogenase (AlaDH) of *Archaeoglobus fulgidus* that catalyzes the NAD$^+$-dependent deamination of ʟ-alanine to ammonium and pyruvate [39]. Thus, HVO_0454 (*ala*) likely catalyzes the first step of ʟ-alanine metabolism (Figure 4) to generate ammonium as a source of fixed nitrogen, as well as pyruvate and reduced cofactor for the biosynthesis of cell carbon and energy.

Based on the transcript profiles, growth on GMM with ʟ-alanine as the N-source also appears to stimulate the production of ʟ-glutamate and acetyl-CoA through uracil and α-ketoglutarate metabolism (Figure 4 and table inset). ʟ-glutamate would be beneficial as an amino-group donor for transaminases that convert α-keto acids to ʟ-amino acids [40], while acetyl-CoA could serve as a substrate for the TCA cycle and an acetyl-group donor for other metabolic reactions [41]. Uracil and α-ketoglutarate are predicted to be converted to ʟ-glutamate and acetyl-CoA during growth on ʟ-alanine based on the following. The transcript levels of the gene neighbors HVO_A0295 (*amaB2*), HVO_A0303 (*dpyS*), HVO_A0305 (*mmsA*), and HVO_A0306 (*gabT6*) were up during grow on ʟ-alanine vs. ammonium. Of these gene products, DpyS and AmaB2 are predicted to convert 5,6-dihydrouracil (an intermediate of uracil metabolism) to β-alanine; DpyS is modeled to be a 3D-structural homolog of a bacterial dihydropyrimidinase [42] and AmaB2 is classified as a β-ureidopropionase/*N*-carbamoyl-ʟ-amino-acid hydrolase in the KEGG database. GabT6 is suggested to transfer the amino group from β-alanine to α-ketoglutarate, as it is modeled to be a 3D-structural homolog of ω-type aminotransferases (that transfer virtually any primary amino group to various ketones [43]) and is divergently transcribed

from the (methyl)malonate-semialdehyde dehydrogenase gene homolog *mmsA* [44]. GabT6 would thus form L-glutamate and malonate-semialdehyde (MSALD), the latter which may be oxidatively decarboxylated by MmsA to generate NADH and acetyl-CoA [44]. The source of uracil to feed into this β-alanine metabolic network may originate from the uracil supplement used to compensate for the Δ*pyrE2* (orotate phosphoribosyl transferase) mutation of the model strain *H. volcanii* H26 [45]. This type of model strain is commonly used in archaeal genetics as a 'wild type' to allow for targeted gene deletion by homologous recombination through uracil selection and 5-fluoroorotic acid (FOA) counterselection.

Table inset associated with figure

Rxn	Gene	Enzyme	Ref
1	*potA1A2BD*, HVO_A0293-4,299-300	Proposed ABC-type alanine transporter	[38]
2	*ala*, HVO_0454	Alanine dehydrogenase [EC 1.4.1.1]	[39]
3	*pykA*, HVO_0806	Pyruvate kinase [EC:2.7.1.40]	[46]
4	*pgk*, HVO_0480	Phosphoglycerate kinase [EC 2.7.2.3]	[47] *
5	*gap*, HVO_0481	Glyceraldehyde phosphate dehydrogenase [EC 1.2.1.12]	[47] *
6	*tpiA*, HVO_2105	Triose phosphate isomerase [EC 5.3.1.1]	[48]
7	*fba*, HVO_1494	Fructose-1,6-bisphosphate aldolase [EC 4.1.2.13]	[47] *
8	*kdgA1*, HVO_0952	2-Keto-3-deoxy-D-phosphogluconate aldolase [EC 4.1.2.14]	[49] *
9	*ppc*, HVO_2621	Phosphoenolpyruvate carboxylase [EC 4.1.1.31]	[50]
10	*fadM2*, HVO_1191	Proline dehydrogenase [EC:1.5.5.2]	[51]
11	*aldH2*, HVO_1189	Proposed delta-1-pyrroline-5-carboxylate dehydrogenase [EC:1.2.1.3]	[52] PDB: 4NMB
12	*uraA*, HVO_1227, HVO_2249, HVO_0335, HVO_B0303	Putative xanthine/uracil permease	[53]
13	Not conserved	5,6-dihydrouracil:NAD(P)+ oxidoreductase [EC:1.3.1.-]; dihydrouracil oxidase [1.3.3.7]	KEGG missing step
14	*dpyS*, HVO_A0303	Proposed dihydropyrimidinase [EC 3.5.2.2]	[42] PDB: 5E5C
15	*amaB2*, HVO_A0295	Proposed β-ureidopropionase / N-carbamoyl-L-amino-acid hydrolase [EC:3.5.1.6 / 3.5.1.87]	KEGG annotation
16	*gabT6*, HVO_A0306	Pyridoxal phosphate-dependent ω-type aminotransferase [EC 2.6.1.19]	[43] PDB: 4A6T
17	*mmsA*, HVO_A0305	Malonate-semialdehyde (MSALD) dehydrogenase [MSDH, EC:1.2.1.27]	[44] IPR010061 family
18	*glpF*, HVO_1542	Putative glycerol facilitator	[54] *
19	*glpK*, HVO_1541	Glycerol kinase [EC 2.7.1.30]	[55] *
20	*glpA1B1C1*, HVO_1538-1540	Glycerol-3-phosphate dehydrogenase [EC 1.1.1.8]	[54] *
21	*amt2*, HVO_0086 *glnK1/2* (PIIs), HVO_0083/85	Amt2 high affinity ammonium transporter and PII (GlnK1/2) regulators (Amt1 was not regulated by log2-fold)	[27] * (PII-Amt binding not function)

Figure 4. Model of metabolic and transport shifts that occur when glycerol grown *H. volcanii* cells use L-alanine instead of ammonium as an N-source. Reactions are defined in the Table inset with those highlighted in green associated with transcripts that are more abundant on L-alanine compared to ammonium. *, reference is specific for *Haloferax* sp. [46–55].

In addition to L-alanine and β-alanine metabolism, genes predicted to function in the oxidation of L-proline to L-glutamate were also upregulated (Figure 4). Included in this list were HVO_1191 (*fadM2* or *putA*), a homolog of the archaeal-type proline dehydrogenase (ProDH) [51], and its gene neighbor HVO_1189 (*aldH2*) which shares predicted 3D structural homology to delta-1-pyrroline-5-carboxylate dehydrogenase (P5CDH) (PDB: 4NMB) [52]. ProDH and P5CDH are enzymes well characterized for their concerted action in converting L-proline to L-glutamate [51,52]. The metabolic signal that would alter the expression of the ProDH and P5CDH gene homologs is unclear. Enhanced levels of L-proline are not predicted to occur in L-alanine vs. ammonium grown cells. Instead, the cells appear to be responding to ammonium limitation based on the enhanced expression of Amt/PII system components on L-alanine vs. ammonium. Thus, the signal for upregulation of ProDH/P5CDH gene homologs could be a general response to ammonium limitation, as the metabolic product of these enzymes, L-glutamate, is central to N-metabolism.

2.4. Identification of a Candidate Regulator and cis-Sequence for Coordinated Transcriptional Control

Genes whose transcript abundance increased significantly on L-alanine compared to ammonium as an N-source clustered into seven distinct regions on the *H. volcanii* genome (Figure 3). The most striking was finding that 11 of the highly expressed genes (HVO_A0293 to HVO_A0306) spanned an 18.6 kb region of plasmid pHV4 at position CP001955.1: 303,472..322,061. The other regions of note were located on the main chromosome, including the *amt2-glnK2 amt1-glnK1* operons at as well as the ProDH (*fadM2*, HVO_1191) and P5CDH (*aldH2*, HVO_1189) homologs at position CP001956.1: 83,382..87,266 and 1,080,889..1,084,135, respectively. Based on this gene clustering, we hypothesized that common promoter elements may coordinate transcriptional responses to the type of N-source.

To identify putative DNA binding motifs that may regulate genes linked to N-shifts, *de novo* motif identification searches were performed (see Materials and Methods for details). Briefly, genes associated with or regulated by the L-alanine N-shift were scanned using MEME for a common motif (input sequences included 5′ regions of *potA1*, *amaB2*, *potD*, *dpyS*, *gabT6*, *ala*, *amt2*, *amt1*, and *lrp*, Table S2). An AT-rich motif with best fit to a semi-palindromic 11 bp sequence AA<u>AGAC</u>T<u>AA</u>RT was identified by this analysis. Using the FIMO program to search the *H. volcanii* genome, this motif was detected in the upstream regions of 492 genes (Table S2), with high statistical support for the motif consensus in regions 5′ of genes differentially expressed in response to L-alanine ($p < 0.0027$, Wilcoxon test vs. randomized sequences, see Methods, Figure 5A, and Table 1). Compared to the rest of the genome, the set of differentially expressed genes was enriched for the motif (hypergeometric test $p < 3.22 \times 10^{-6}$, Table S2). Eleven of the 30 differentially expressed genes were located within the 16-gene cluster involved in N-transport and metabolism, with the high confidence motifs located upstream of *potA1*, *potD*, *mmsA*, and *gabT6* (all highly expressed on L-alanine) (Figures 2A and 5B, Table 1).

Of the four genes identified to have high confidence AT-rich motifs, *mmsA* and *gabT6* were found to be in genome synteny with the transcription factor homolog HVO_A0307 (*lrp*) (Figures 3 and 5). This type of genomic organization is observed in other archaea [56]. Of particular note is the *Sulfolobus* BarR (Saci_2136 and ST1115) Lrp-type transcription factor that shares 32% identity with HVO_0307 (*lrp*) and is genetically linked to *mmsA* and/or *gabT6* gene homologs [56,57], including the *Sulfolobus* ST1116 that is 45% identical to *mmsA* and the *Sulfolobus* ω-type aminotransferases Saci_2137 and ST1114 that are 43%-44% identical to *gabT6*. The *Sulfolobus* BarR binds a semi-palindromic AT-rich motif that repeats evenly in an intergenic region 5′ of the divergently transcribed *barR* and the *gabT6*-like gene [56]. Similarly to *Sulfolobus* BarR, HVO_A0307 is of the full-length feast/famine regulatory protein (FFRP) subgroup of the Lrp family transcriptional regulators that have an N-terminal DNA binding domain and a C-terminal domain that promotes self-assembly [58]. When compared to the X-ray crystal structures in the PDB database (August 30, 2019) by Phyre2-based structural homology modeling, HVO_A0307 and BarR were found to be most closely related to the *Sulfolobus tokodaii St*Grp (Lrp-type glutamine receptor protein) (PDB: 2E7W) (Figure 6). Further analysis revealed HVO_A0307 (*lrp*) to

have conserved amino acid residues with BarR and *St*Grp at positions proposed to interact with ligand and/or influence self-assembly of the transcription factor [58]. Thus, HVO_A0307 is proposed to be an Lrp-type transcriptional regulator that may bind the repetitive AT-rich motifs 5′ of *potA1*, *potD*, *mmsA*, and *gabT6* (Figure 5B) in response to metabolic intermediates that signal growth on ʟ-alanine vs. ammonium as the N-source (e.g., ʟ-alanine, β-alanine, ʟ-glutamine, and/or ʟ-glutamate). The genes with this 5′ AT-rich motif (*potA1*, *potD*, *mmsA*, and *gabT6*) are in apparent operons with the ʟ-alanine induced *potA2*, *amaB2*, *hvo_a0295a*, *hvo_a0295*, *hvo_a0296*, *potB,* and *hvo_a0301,* consistent with the hypothesis that the entire ʟ-alanine-induced pHV4 region is regulated by the Lrp-related HVO_A0307.
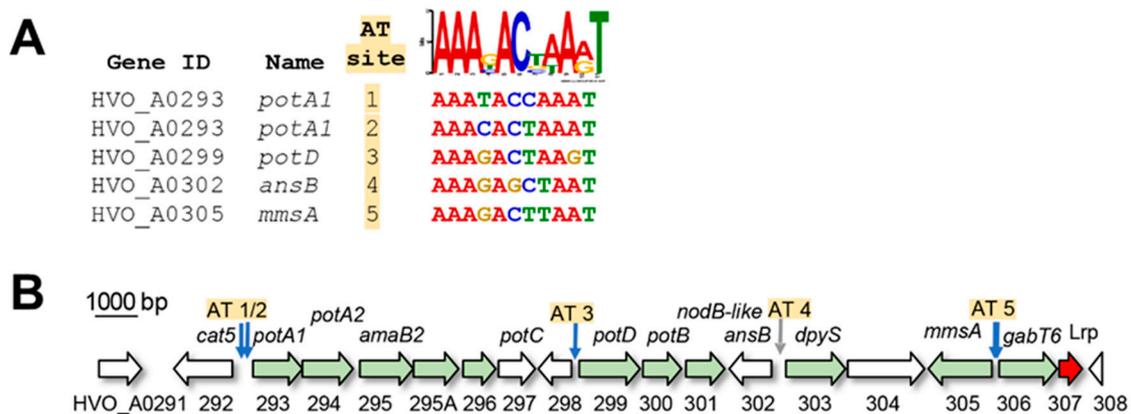
Figure 5. Putative cis-regulatory sequence motif and Lrp-like transcriptional regulator located in the 11 gene N-transport and metabolism cluster. (**A**) Motif logo in upper right corner depicts the position weight matrix results from the computational motif search. Genes containing the motif are listed at left, with motif sequences identified by genome scanning (FIMO) listed at right under the motif logo. AT-rich site numbers correspond to genomic locations within the gene cluster shown in (B). (**B**) Genomic location of motifs (vertical arrows) relative to differentially expressed ORFs (green horizontal arrows) within the HVO_A0291 – HVO_A0307 cluster. Putative Lrp-like transcriptional regulator is indicated by the red arrow. Blue vertical arrows indicate high confidence motifs, and the grey arrow indicates a low confidence motif (see also Table S2 and Figure A3).
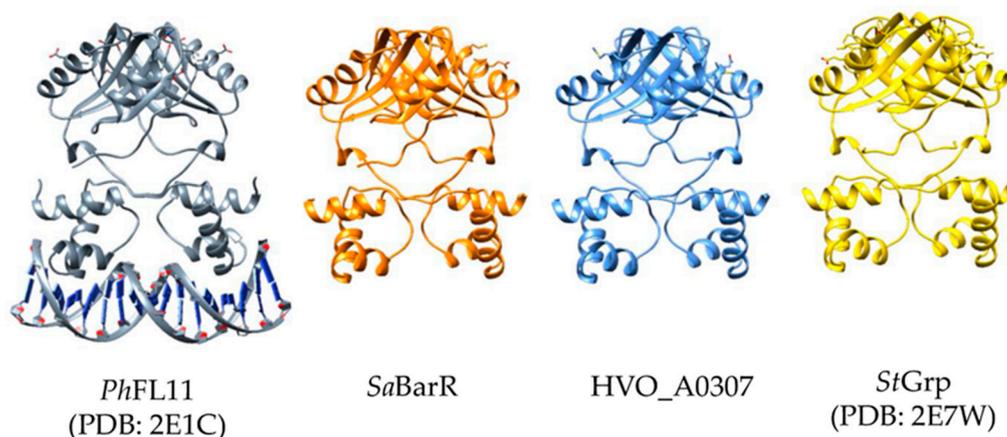
Figure 6. *Cont.*

Lrp	Ligand	Residues associated with ligand binding and subunit-subunit interactions										
*Ph*FL11 (PDB:2E1C)	K>>R>>Q	Y100	E118	T120	D124	L141	D142	G145	T153	H154	T155	L159
*Ph*DM1 (PDB:2Z4P)	I>V>R, L, M>F	E15	V33	Y35	D39	I56	T57	R61	T69	S70	T71	I75
HVO_A0307 (3D model)	?	Y87	Y106	M108	D112	I129	D130	V133	T141	S142	S143	M147
*Sa*BarR (3D model)	β-Ala	Y101	F120	L122	D126	L144	E145	M148	T156	S157	S158	V162
*St*Grp (PDB: 2E7W)	Q	Y77	F96	L98	D102	L120	E121	M124	T132	S133	T134	V138

Figure 6. Full-length FFRP subgroup member HVO_A0307 and its relationship to other Lrp-type regulators. Upper: X-ray crystal structures of *Ph*FL11 (PDB: 2E1C) bound to DNA and *St*Grp (PDB: 2E7W) compared to 3D models of *Sa*BarR (Saci_2136) and HVO_A0307. When compared to all X-ray crystal structures available in the PDB databases (August 30, 2019) by homology modeling, HVO_A0307 was most closely related to *St*Grp (Lrp-type glutamine receptor protein) (PDB: 2E7W) at an root mean square deviation (RMSD) between 134 pruned atom pairs of 0.576 Å; (across all 149 pairs: 1.234 Å). The 3D models of HVO_A0307 and *Sa*BarR have an RMSD between 140 pruned atom pairs of 0.240 Å; (across all 148 pairs: 0.884 Å). Table inset: Residues of Lrp-type proteins found at the positions at which *Ph*DM1 interacts with ligand and/or alters subunit interactions [58]. Identical and similar residues to HVO_A0307 are highlighted in green. Abbreviations: Ph, *Pyrococcus horikoshii*; Sa, *Sulfolobus acidocaldarius*; St, *Sulfolobus tokodaii*.

3. Materials and Methods

3.1. Growth of H. volcanii

H. volcanii H26 was grown at 42 °C with shaking (200 rpm) in ATCC 974 complex medium (Hv-CM) or glycerol minimal medium (GMM). GMM was as previously described [45,59] with 20 mM glycerol as the main carbon source and 10 mM ammonium chloride (NH_4Cl) or L-alanine as the N-source; uracil (50 μg·ml^{-1}) was included in GMM to allow for growth of H26 Δ*pyrE2* [45,59]. The GMM formula per liter was: 20 mM glycerol, 141 g NaCl, 17.6 g $MgCl_2$, 20.6 g $MgSO_4$, 4.12 g KCl, 432 mg $CaCl_2$, 0.353 mg $MnCl_2$, 0.432 mg $ZnSO_4$, 2.26 mg $FeSO_4$, 0.0471 mg $CuSO_4$, 50 mg uracil, 0.882 mg biotin, 0.882 mg thiamine, 1.91 mL of 0.5 M KPO_4 buffer (pH7.5), and 35.3 mL of 1 M Tris-Cl (pH7.5) buffer with 10 mM NH_4Cl or L-alanine as the variable.

A previous method was used to monitor growth rates under these conditions [60]. Briefly, single colonies of H26 were first inoculated in 5-mL Hv-CM and grown aerobically at 42 °C to early stationary phase (OD_{600nm}, ~1.0). Cells were harvested via centrifugation (15,871× *g*, 1 min at room temperature) and washed two times with GMM without added N-sources. Cultures were then diluted to a starting OD_{600nm} of 0.03 in 200 μL of GMM with 10 mM NH_4Cl or L-alanine under continuous shaking at 42 °C in a Bioscreen C analysis system (Growth Curves USA, Piscataway, NJ, USA) set to measure OD_{600nm} every 30 min. Each condition was tested using four independent biological replicate samples, each with three technical replicates.

3.2. Preparation and Analysis of Microarray Data

H26 was grown aerobically to exponential phase (OD_{600nm}, 0.3 to 0.5) in GMM with two different N-sources (10 mM), NH_4Cl or L-alanine. Total RNA preparation, NimbleGen microarray slides, cDNA synthesis and dye hybridization were as previously reported [61]. Double-stranded cDNA libraries were generated using the Superscript cDNA synthesis kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. One microgram of the library from each biological replicate was labeled with Cy3 dye and hybridized to NimbleGen 12 × 135-k feature single-color custom microarray slides as described in the kit (Roche NimbleGen, Inc., Madison, WI, USA), with each 135-k array containing 98%

of the annotated genes in the *H. volcanii* genome (3,985 genes, NCBI Genome ID, 1149) [61]. Microarray hybridization and scanning were conducted at the FSU-NimbleGen-certified facility (The Florida State University, Tallahassee, FL, USA). For each gene, 96 replicate data points were measured (32 replicate probes per gene per array, with 3 biological replicate hybridizations per sample). Raw spot intensities were first normalized within arrays using RMA, followed by normalization and analysis using the Subio Platform v. 1.22 (https://www.subioplatform.com) with the following parameters: filtering out signals below the low signal cutoff (raw intensity <1.0), global normalization (75th percentile) and log2 transformation. Subsequent statistical tests, plotting, and analysis of microarray data were conducted R statistical computing environment (http://www.R-project.org). Resultant expression intensities (Table S1) were averaged across the four replicate probes for each gene in each growth condition, then subject to Student's *t* test followed by Benjamini-Hochberg correction for multiple hypothesis testing [62]. The *q* value (false discovery rate, FDR) <0.05 was set as a cutoff, yielding the final list of 30 genes in Table 1. To generate heatmaps, genes were clustered by gene expression pattern across the three biological replicates using average linkage hierarchical clustering and plotted using the ggplot2 package [63]. Significance of enrichment of the differentially expressed genes by functional categories was determined using the hypergeometric test with Benjamini-Hochberg correction. Annotations were computed using eggNOG-mapper [32] based on eggNOG 4.5 orthology data. eggNOG is a public database of orthologous groups across different taxonomic levels and leverages several databases and text mining to call functional predictions. All code used in this study can be found in the GitHub repository: https://github.com/sungminhwang-duke/Microarray_N_sources. The microarray platform used in this study is available through NCBI Gene Expression Omnibus (GEO) at accession number GPL21414, and the raw and processed microarray data are available at accession number GSE130934.

3.3. Computational Prediction of Transcription Factor Binding Sequences

Computational prediction of DNA motifs was performed as follows. DNA sequences in the 5' direction of selected genes differentially expressed in response to nitrogen or within operons of differentially expressed genes (*potA1*, *amaB2*, *potD*, *dpyS*, *gabT6*, *ala*, *amt2*, *amt1*, and *lrp*, in Table S2) were retrieved using the graphics tool within the NCBI nucleotide portal (https://www.ncbi.nlm.nih.gov/nuccore/). De novo motif detection was performed with these primary DNA sequences as input using the MEME Suite version 5.0.5 [64]. The parameters were set to any number of repeats, maximum width of 14-17 bp, reverse complement motifs allowed, and 3 output motifs. The FIMO algorithm from the MEME Suite was used to scan the *H. volcanii* genome (uid46845 June 11,2018 version within the Upstream Sequences: Prokaryotic database) for additional instances of the putative motifs using default parameters. Sequences containing the motif most often associated with differentially expressed genes were shuffled and compared to original sequences using the Wilcoxon signed-rank test to determine significance. Shuffling was conducted to preserve dinucleotide frequencies using the fasta-shuffle-letters command within the MEME suite (the underlying algorithm was based on uShuffle [65]). Details regarding the DNA motif analysis (input sequences, FIMO output) are given in Table S2.

4. Conclusions

Global transcript profiling of *H. volcanii* grown on a C3 source of glycerol with abundant (ammonium) verses intermediate (L-alanine) sources of fixed nitrogen revealed a distinct set of genes that were regulated including 20 genes that were upregulated and 10 genes that were downregulated on L-alanine. Of the upregulated genes, the majority could be mapped to the transport (7 genes) and metabolism (7 genes) of ammonium, L-alanine, and/or other associated metabolites. The transport systems included the high affinity ammonium transporter homolog Amt2 and its PII regulators GlnK1/2 as well as the ABC-type PotA1A2BD system which is suggested to transport amino acids such as L-alanine. Based on the metabolic genes that were upregulated, the oxidative deamination of L-alanine to pyruvate appeared to be of central importance to the L-alanine grown cells. Gene

neighbors encoding an apparent pathway to synthesize ʟ-glutamate and acetyl-CoA from uracil and α-ketoglutarate also were enhanced at the level of transcript abundance under the ʟ-alanine conditions. Genome clustering of the upregulated genes was observed. This clustering enabled us to generate a model that the Lrp-type HVO_A0307 transcription factor homolog is a candidate for regulating the transcription of genes associated with N-shifts from ammonium to ʟ-alanine. Thus, we propose that transcriptional co-regulation of a syntenic cluster of operons enables *H. volcanii* to respond to and utilize intermediate vs. abundant sources of fixed nitrogen for growth.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/19/4784/s1.

Author Contributions: All authors contributed to the conceptualization, investigation, methodology and review/editing of the paper. Other author contributions included: microarray sample preparation (N.E.C.); software (S.H., R.K.H. and A.K.S.); validation, formal analysis, and writing—original draft preparation (S.H., A.K.S. and J.A.M.-F.); and resources, supervision and funding acquisition (A.K.S. and J.A.M.-F.).

Conflicts of Interest: The authors declare no conflict of interest.

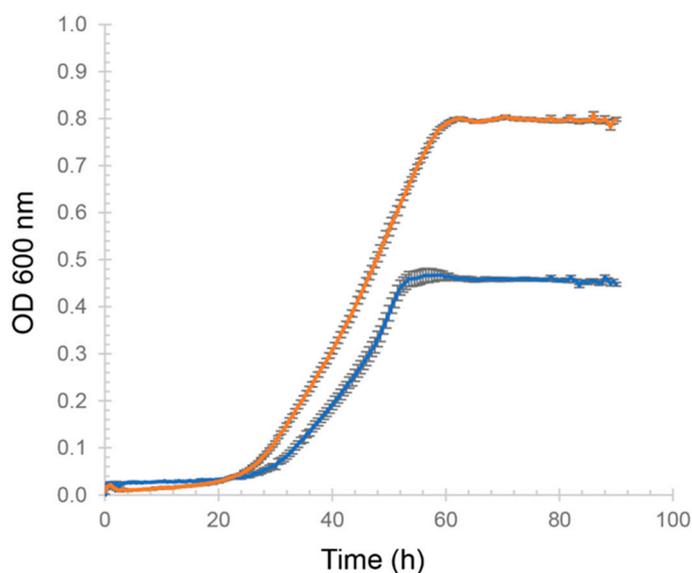## Appendix A Supplemental Figures

Figure A1. Growth of *H. volcanii* H26 in minimal medium with the N-source of ammonium (blue) and ʟ-alanine (orange). Total *n* = 12 (biological quadruplicate and technical triplicate) was applied. Error bars, standard deviation of the mean. See Supplemental Methods in Appendix B for details.
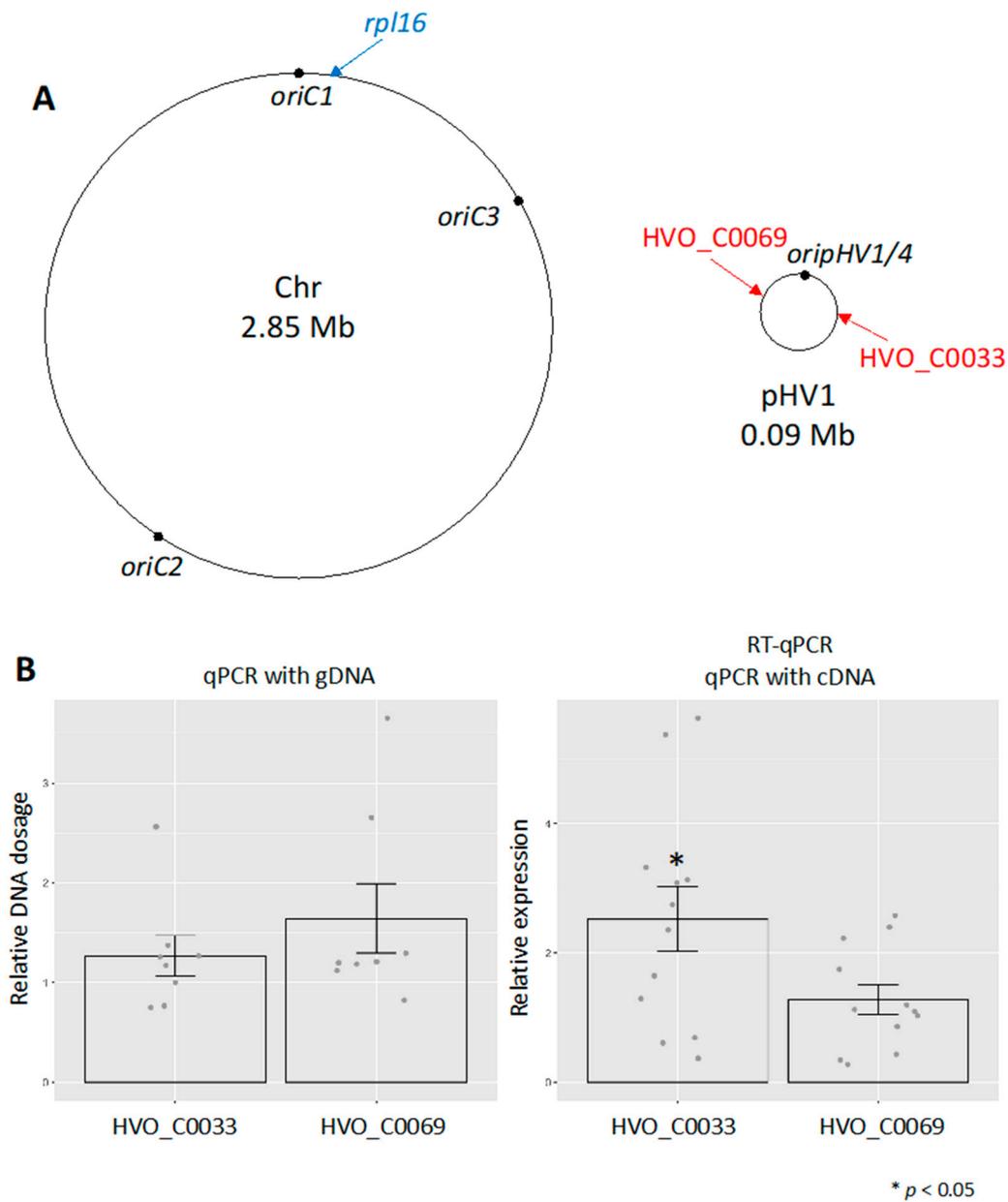
Figure A2. (**A**) Genomic locations of amplification targets for qPCR. Target genes in main chromosome and pHV1 are indicated in blue and red, respectively. Black dots show origin of replication in both chromosome and pHV1. (**B**) Relative copy number with gDNA (left) and cDNA (right) templates. Asterisks indicate $p < 0.05$ comparing the reference gene *rpl16* to sites of interest on pHV1 (See Supplemental Methods in Appendix B for details). Relative DNA dosage and mRNA gene expression (y-axes) are ratios of templates extracted from ammonium- vs. L-alanine-grown cells. Total n = 12 (biological quadruplicate and technical triplicate) and *n* = 8 (biological quadruplicate and technical duplicate) were applied to qPCR with cDNA and gDNA, respectively.

```
HVO_A0292< CATGCCTTTCGTTCCGAAAGTTCGCATCTTAACTATTGCGCTCTTACGAGCGTTCTCCGCGGATCACGCGGTACCCGACCACGTGTCGGACGCTTCTCGCCGAGCGCGTACGCTTGATTCGA
GCCGAGCGCCGCGGTCTCCTGCCAACACGTACGGACACACATTCCCAGTCTGAACGGATTAGGAAAAATTCCACATAAACGAGTAGCTAGTGGTCAGAAATAGATTAGAAACCGATATAGATA AAACAATATT
                                                                                                              AT1
CAGCTAACTGGCCTACTATACACAATTTGTCCGGCA AAATACCAAAT TACAGCGTTATAATTGTAACATCTGAACCTCGTCTGCTCGGCCGCCCCGGTGAGTCGGTCCGACCGGAGTATCGGCGGTGGAG
                                                                        AT2
TCGGGTCGTATGAGTTCTCACCCGCGCGTAGCGTGCGAGTCTGTGTGAC AAAGA AAACACTAAAT GCACGCATGTAACATATGATGGAT ATG >HVO_A0293 potA1

HVO_A0298< CACGTGGTGTCAATTCATATCGCGGTACAAGTAGTTAGCGACGGGGGAGCAACCACGGGAGCGAAACGTCCGTTTCGGTCCGCTCCGTCGGGCGGGCGGGTGAAACCCAATCGGTCACACAAA
                              AT3
TATCCGAAAAAGACGC AAAGACTAAGT AGTGTCGTTATAGACTCCTATTCCATGCCTGATAGC ATG >HVO_A0299 potD
                                                              AT4 (−)
ansB HVO_A0302< CATATGAGTGCTGTTGGGGCTTCCGCGG ATTAGCTCTTT CGTCCACGCAAACGCGAGCGGCGGCCCGGTGAGACCCCGCGACCACGGCCGGCGACGCTCGACCCGCCGACGACCGCCGA
CACACCTCAACGGCGAACGTGTCGGTCGAGGGAACAATGCAGCAGTCTGCCAAACAATATTAAGAATCAATGGAACTATTACGGATTATAGCTCTGTAATTTGGAGGGCGGTGACTCCGGTTGCAC CGCCTGGTG
ACGGACGAAACGAATCTGACGAACGACTCGGTTCAGCCGGCGATCGGGTGAACCTTTTTACTATCGGCCCGGAATTATCGATGGT ATG >HVO_A0303 dpyS
                                                              AT5 (−)
mmsA HVO_A0305< CATGGTACGTTCGACTGCAGTTCGCTCCCA ATTAAGTCTTT CCTACGATTTTCGAAATATAGCTCT AAAGAAAGAATTTGTCCATTGAATTTC AAAGA GTCTCGAACGCTGTTCTAGA
AATAAGCATCTAGTGC AAAGA TTAATCAATAGCAGGGTAGTAGTCGTTCAACAACG ATG >HVO_A0306 gabT6
                                                              AT6
HVO_C0050< CATTATCTGTCGGTTCAGACCACAACACTATATCTCTTCTGAACCACTTTGGTTTATTCAAT AAATACCAACT TGGGTGATCCGAACTGTTATAGCGCCGCTGCTCGTAGAGACACGT ATG >HVO_C0051
```

Figure A3. Sequences of putative regulatory motifs detected upstream of genes differentially expressed during a N-shift from ammonium to L-alanine (see also Figure 5 in main text). See Materials and Methods for motif detection details. AT-rich sites detected by the FIMO program with *p*-value $< 1 \times 10^{-7}$ are shown in yellow (high confidence sites labelled AT1, 2, 3, 5) and those with *p*-value $> 1 \times 10^{-7}$ in blue (low confidence sites labelled AT4, 6). AAAGA repeats outside of the AT-rich motif are highlighted in blue. Start codons are underlined. (−) represents DNA motif of complementary DNA sequence.

Appendix B Supplemental Methods

Appendix B.1 Growth Assay

Growth assay was as previously described [60]. Briefly, single colonies of H26 were inoculated in 5-mL Hv-CA (casamino acid) and grown aerobically at 42 °C to early stationary phase to synchronize growth phase (OD$_{600nm}$, ~1.0). Cells were harvested via centrifugation (15,871× *g*, 1 min at room temperature) and washed two times with minimal medium without added N-sources. Cultures were diluted to a starting OD$_{600nm}$ of 0.03 in 200 μL of minimal medium containing uracil with 10 mM NH$_4$Cl or L-alanine under continuous shaking at 42 °C in a Bioscreen C analysis system (Growth Curves USA, Piscataway, NJ, USA) set to measure every 30 min at OD$_{600nm}$. Each condition was tested using four biological samples, each with three technical replicates, as presented in Figure A1.

Appendix B.2 Quantitative PCR to Calculate Copy Number and Gene Expression

Genomic DNA (gDNA) was extracted from 3 mL of cell culture as described by Dulmage et al. [33]. In brief, four biological replicate H26 were grown in minimal medium (see main text methods for details on media formulations), including uracil (50 μg·mL^{-1}), with 10 mM L-alanine or 10 mM NH$_4$Cl. Cells were harvested at OD$_{600nm}$ ~ 0.5 by centrifugation (15,871× *g*, 30 s). The cell pellet was added to 250 μL of water and 250 μL of lysis buffer with the recipe described in *The Halohandbook* [59] and incubated for 5 min at room temperature (RT) with an occasional vortex. RNA was removed by RNase A (100 μg, 5 min at RT) and protein was degraded by Proteinase K (100 μg) at 37 °C for 10 min. Cell lysates were mixed with an equal volume of phenol/chloroform/ isoamyl alcohol (25:24:1) and DNA was precipitated in ethanol and resuspended in TE buffer. In addition, total RNA was isolated by using an Absolute RNA Miniprep kit (Agilent Technologies, Santa Clara, CA, USA) according to the procedures provided in the kit. An additional DNA digestion step was applied by using a Turbo DNA-free kit (Invitrogen, Carlsbad, CA, USA). The level of DNA contamination after the DNase treatment was checked by 35-cycles of end-point PCR with primers (HVO0863external_FW and HVO0863external_RV) targeting up/downstream of HVO_0863 (see primers in Table A1). For quantification, ten nanograms of gDNA or total RNA per reaction mixture volume (10 μL) served as the template. Quantitative PCR (qPCR) was carried out by qPCRBIO SyGreen 1-Step Go Lo-R kit (PCR Biosystems Inc., Wayne, PA, USA). One-step qPCR was performed under conditions of 45 °C for 10 min, 95 °C for 2 min, and 40 cycles of 95 °C for 15 s and 60 °C for 30 s, followed by determination of the melting curve by using a LightCycler 96 (Roche, Madison, WI, USA). The expression level of HVO_C0033 and HVO_C0069 were normalized

to the internal gene, *rpl16*. The relative gene expression in ammonium-grown vs. ʟ-alanine-grown cells was calculated by Livak's method. Oligonucleotides used for qPCR (Table A1) were quantified for efficiency, which was determined to be between 90 and 110 percent. Gene expression for each target gene was compared to the internal *rpl16* gene to calculate significance by Welch's *t*-test, and the correlation of the two independent experiments, microarray and RT-qPCR, was calculated by Spearman's method.

Table A1. Oligonucleotides for quantitative PCR in this study.

Name	Sequence (5′-3′)	Description	Ref.
HVO0863external_FW	TCCCGGAGAGGTGTGTATGT	To test DNA contamination. Targeting up/downstream of HVO_0863. Amplicon: 1,767 bp.	This study
HVO0863external_RV	TCACTGACCTCGAACACGTC	To test DNA contamination. Targeting up/downstream of HVO_0863. Amplicon: 1,767 bp.	This study
rpl16s1_F	CCACGTCATCCGCGAGAACA	Control locus in main chromosome. Amplicon: 86 bp.	[66]
rpl16s1_R	CGACCTTCCCGAACGACTGG	Control locus in main chromosome. Amplicon: 86 bp.	[66]
qHVO_C0033_s2_F	CTCGGTTCGACAGGCCAAGG	Target locus in pHV1. Amplicon: 113 bp.	This study
qHVO_C0033_s2_R	CAGTTCCGCACATGGCATCG	Target locus in pHV1. Amplicon: 113 bp.	This study
qHVO_C0069_s2_F	AGGCACCGACTGTCTCGTCA	Target locus in pHV1. Amplicon: 124 bp.	This study
qHVO_C0069_s2_R	GAAGGAGAATCGCGCGGTGA	Target locus in pHV1. Amplicon: 124 bp.	This study
qHVO_A0299_s2_F	ACCGCCTACGGAACCTCAGT	Target locus in pHV4. Amplicon: 126 bp.	This study
qHVO_A0299_s2_R	CTCGGACATCGTGGCGTAGG	Target locus in pHV4. Amplicon: 126 bp.	This study
qHVO_0454_s2_F	CGGAGCTCATCTCGGCCATC	Target locus in main chromosome. Amplicon: 92 bp.	This study
qHVO_0454_s2_R	CTGCGGCAGGTCGATGTAGG	Target locus in main chromosome. Amplicon: 92 bp.	This study
qHVO_0086_s1_F	CGCGTGGATTGACTGGCTCT	Target locus in main chromosome. Amplicon: 100 bp.	This study
qHVO_0086_s1_R	GATGTACGCGCGGAAGTCCA	Target locus in main chromosome. Amplicon: 100 bp.	This study
qHVO_1189_s1_F	CTCCGAGGACTACCGCCAGA	Target locus in main chromosome. Amplicon: 81 bp.	This study
qHVO_1189_s1_R	CCGATACACGGGAGGTTGCC	Target locus in main chromosome. Amplicon: 81 bp.	This study
qHVO_2307_s1_F	TTCGACCTCGACGGAACCCT	Target locus in main chromosome. Amplicon: 127 bp.	This study
qHVO_2307_s1_R	CGAGTCCGTGCCAGATGGTC	Target locus in main chromosome. Amplicon: 127 bp.	This study
qHVO_C0064_s1_F	GCGAATACGTCGTCCACGGT	Target locus in pHV1. Amplicon: 70 bp.	This study
qHVO_C0064_s1_R	TGGCTCTCGCAGCTGTTGAC	Target locus in pHV1. Amplicon: 70 bp.	This study

References

1. Alwine, J.C.; Kemp, D.J.; Stark, G.R. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5350–5354. [CrossRef] [PubMed]

2. Wang, A.M.; Doyle, M.V.; Mark, D.F. Quantitation of mRNA by the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9717–9721. [CrossRef] [PubMed]

3. Selinger, D.W.; Cheung, K.J.; Mei, R.; Johansson, E.M.; Richmond, C.S.; Blattner, F.R.; Lockhart, D.J.; Church, G.M. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat. Biotechnol.* **2000**, *18*, 1262–1268. [CrossRef] [PubMed]

4. Lashkari, D.A.; DeRisi, J.L.; McCusker, J.H.; Namath, A.F.; Gentile, C.; Hwang, S.Y.; Brown, P.O.; Davis, R.W. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 13057–13062. [CrossRef] [PubMed]

5. Nagalakshmi, U.; Wang, Z.; Waern, K.; Shou, C.; Raha, D.; Gerstein, M.; Snyder, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **2008**, *320*, 1344–1349. [CrossRef] [PubMed]

6. Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628. [CrossRef]

7. Danan, M.; Schwartz, S.; Edelheit, S.; Sorek, R. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* **2012**, *40*, 3131–3142. [CrossRef] [PubMed]

8. Esclapez, J.; Pire, C.; Camacho, M.; Bautista, V.; Martinez-Espinosa, R.M.; Zafrilla, B.; Vegara, A.; Alcaraz, L.A.; Bonete, M.J. Transcriptional profiles of *Haloferax mediterranei* based on nitrogen availability. *J. Biotechnol.* **2015**, *193*, 100–107. [CrossRef] [PubMed]

9. Twellmeyer, J.; Wende, A.; Wolfertz, J.; Pfeiffer, F.; Panhuysen, M.; Zaigler, A.; Soppa, J.; Welzl, G.; Oesterhelt, D. Microarray analysis in the archaeon *Halobacterium salinarum* strain R1. *PLoS ONE* **2007**, *2*, e1064. [CrossRef]

10. Schmid, A.K.; Reiss, D.J.; Pan, M.; Koide, T.; Baliga, N.S. A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability. *Mol. Syst. Biol.* **2009**, *5*, 282. [CrossRef]

11. Richardson, D.J.; Watmough, N.J. Inorganic nitrogen metabolism in bacteria. *Curr. Opin. Chem. Biol.* **1999**, *3*, 207–219. [CrossRef]

12. Stahl, D.A.; de la Torre, J.R. Physiology and diversity of ammonia-oxidizing archaea. *Annu. Rev. Microbiol.* **2012**, *66*, 83–101. [CrossRef]

13. Kowalchuk, G.A.; Stephen, J.R. Ammonia-oxidizing bacteria: A model for molecular microbial ecology. *Annu. Rev. Microbiol.* **2001**, *55*, 485–529. [CrossRef]

14. van Spanning, R.J.M.; Richardson, D.J.; Ferguson, S.J. *Introduction to the Biochemistry and Molecular Biology of Denitrification*; Elsevier B.V.: Amsterdam, The Netherlands, 2007.

15. Bae, H.S.; Morrison, E.; Chanton, J.P.; Ogram, A. Methanogens are major contributors to nitrogen fixation in soils of the Florida Everglades. *Appl. Environ. Microbiol.* **2018**, *84*, e02222-17. [CrossRef] [PubMed]

16. Hoffman, B.M.; Lukoyanov, D.; Yang, Z.Y.; Dean, D.R.; Seefeldt, L.C. Mechanism of nitrogen fixation by nitrogenase: The next stage. *Chem. Rev.* **2014**, *114*, 4041–4062. [CrossRef]

17. Lam, P.; Kuypers, M.M. Microbial nitrogen cycling processes in oxygen minimum zones. *Ann. Rev. Mar. Sci.* **2011**, *3*, 317–345. [CrossRef] [PubMed]

18. Ward, M.H.; Jones, R.R.; Brender, J.D.; de Kok, T.M.; Weyer, P.J.; Nolan, B.T.; Villanueva, C.M.; van Breda, S.G. Drinking water nitrate and human health: An updated review. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1557. [CrossRef]

19. Griffis, T.J.; Chen, Z.; Baker, J.M.; Wood, J.D.; Millet, D.B.; Lee, X.; Venterea, R.T.; Turner, P.A. Nitrous oxide emissions are enhanced in a warmer and wetter world. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12081–12085. [CrossRef] [PubMed]

20. Vassart, A.; Van Wolferen, M.; Orell, A.; Hong, Y.; Peeters, E.; Albers, S.V.; Charlier, D. Sa-Lrp from *Sulfolobus acidocaldarius* is a versatile, glutamine-responsive, and architectural transcriptional regulator. *Microbiologyopen* **2013**, *2*, 75–93. [CrossRef]

21. Jager, D.; Sharma, C.M.; Thomsen, J.; Ehlers, C.; Vogel, J.; Schmitz, R.A. Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21878–21882. [CrossRef]

22. Prasse, D.; Schmitz, R.A. Small RNAs involved in regulation of nitrogen metabolism. *Microbiol. Spectr.* **2018**, *6*. [CrossRef]

23. Esclapez, J.; Bravo-Barrales, G.; Bautista, V.; Pire, C.; Camacho, M.; Bonete, M.J. Effects of nitrogen sources on the nitrate assimilation in *Haloferax mediterranei*: Growth kinetics and transcriptomic analysis. *FEMS Microbiol. Lett.* **2014**, *350*, 168–174. [CrossRef] [PubMed]

24. Paya, G.; Bautista, V.; Camacho, M.; Castejon-Fernandez, N.; Alcaraz, L.A.; Bonete, M.J.; Esclapez, J. Small RNAs of *Haloferax mediterranei*: Identification and potential involvement in nitrogen metabolism. *Genes* **2018**, *9*, 83. [CrossRef] [PubMed]

25. Mancinelli, R.L.; Hochstein, L.I. The occurrence of denitrification in extremely halophilic bacteria. *FEMS Microbiol. Lett.* **1986**, *35*, 55–58. [CrossRef] [PubMed]

26. Martinez-Espinosa, R.M.; Marhuenda-Egea, F.C.; Bonete, M.J. Purification and characterisation of a possible assimilatory nitrite reductase from the halophile archaeon *Haloferax mediterranei*. *FEMS Microbiol. Lett.* **2001**, *196*, 113–118. [CrossRef]

27. Pedro-Roig, L.; Lange, C.; Bonete, M.J.; Soppa, J.; Maupin-Furlow, J. Nitrogen regulation of protein-protein interactions and transcript levels of GlnK PII regulator and AmtB ammonium transporter homologs in Archaea. *MicrobiologyOpen* **2013**, *2*, 826–840. [CrossRef] [PubMed]

28. Mullakhanbhai, M.F.; Larsen, H. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* **1975**, *104*, 207–214. [CrossRef] [PubMed]

29. Stiller, M.; Nissenbaum, A. Geochemical investigation of phosphorus and nitrogen in the hypersaline Dead Sea. *Geochim. Cosmochim. Acta* **1999**, *63*, 3467–3475. [CrossRef]

30. Large, A.; Stamme, C.; Lange, C.; Duan, Z.; Allers, T.; Soppa, J.; Lund, P.A. Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential *cct1* gene. *Mol. Microbiol.* **2007**, *66*, 1092–1106. [CrossRef] [PubMed]

31. Lie, T.J.; Leigh, J.A. Regulatory response of *Methanococcus maripaludis* to alanine, an intermediate nitrogen source. *J. Bacteriol.* **2002**, *184*, 5301–5306. [CrossRef]

32. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **2016**, *44*, D286–D293. [CrossRef] [PubMed]

33. Dulmage, K.A.; Darnell, C.L.; Vreugdenhil, A.; Schmid, A.K. Copy number variation is associated with gene expression change in archaea. *Microb. Genom.* **2018**, *4*, e000210. [CrossRef]

34. Maurer, S.; Ludt, K.; Soppa, J. Characterization of copy number control of two *Haloferax volcanii* replication origins using deletion mutants and haloarchaeal artificial chromosomes. *J. Bacteriol.* **2018**, *200*, e00517-17. [CrossRef] [PubMed]

35. Ausiannikava, D.; Mitchell, L.; Marriott, H.; Smith, V.; Hawkins, M.; Makarova, K.S.; Koonin, E.V.; Nieduszynski, C.A.; Allers, T. Evolution of genome architecture in archaea: Spontaneous generation of a new chromosome in *Haloferax volcanii*. *Mol. Biol. Evol.* **2018**, *35*, 1855–1868. [CrossRef] [PubMed]

36. Zerulla, K.; Baumann, A.; Soppa, J. *Cell Cycle and Polyploidy in Haloarchaea*; Caister Academic Press: Wymondham, UK, 2014; pp. 145–166.

37. Conroy, M.J.; Durand, A.; Lupo, D.; Li, X.D.; Bullough, P.A.; Winkler, F.K.; Merrick, M. The crystal structure of the *Escherichia coli* AmtB-GlnK complex reveals how GlnK regulates the ammonia channel. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1213–1218. [CrossRef] [PubMed]

38. Kadaba, N.S.; Kaiser, J.T.; Johnson, E.; Lee, A.; Rees, D.C. The high-affinity *E. coli* methionine ABC transporter: Structure and allosteric regulation. *Science* **2008**, *321*, 250–253. [CrossRef] [PubMed]

39. Schroder, I.; Vadas, A.; Johnson, E.; Lim, S.; Monbouquette, H.G. A novel archaeal alanine dehydrogenase homologous to ornithine cyclodeaminase and mu-crystallin. *J. Bacteriol.* **2004**, *186*, 7680–7689. [CrossRef] [PubMed]

40. Walker, M.C.; van der Donk, W.A. The many roles of glutamate in metabolism. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 419–430. [CrossRef]

41. Shi, L.; Tu, B.P. Acetyl-CoA and the regulation of metabolism: Mechanisms and consequences. *Curr. Opin. Cell Biol.* **2015**, *33*, 125–131. [CrossRef]

42. Tzeng, C.T.; Huang, Y.H.; Huang, C.Y. Crystal structure of dihydropyrimidinase from *Pseudomonas aeruginosa* PAO1: Insights into the molecular basis of formation of a dimer. *Biochem. Biophys. Res. Commun.* **2016**, *478*, 1449–1455. [CrossRef]

43. Humble, M.S.; Cassimjee, K.E.; Hakansson, M.; Kimbung, Y.R.; Walse, B.; Abedi, V.; Federsel, H.J.; Berglund, P.; Logan, D.T. Crystal structures of the *Chromobacterium violaceum* ω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP. *FEBS J.* **2012**, *279*, 779–792. [CrossRef] [PubMed]

44. Stines-Chaumeil, C.; Talfournier, F.; Branlant, G. Mechanistic characterization of the MSDH (methylmalonate semialdehyde dehydrogenase) from *Bacillus subtilis*. *Biochem. J.* **2006**, *395*, 107–115. [CrossRef] [PubMed]

45. Allers, T.; Ngo, H.P.; Mevarech, M.; Lloyd, R.G. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* **2004**, *70*, 943–953. [CrossRef] [PubMed]

46. Johnsen, U.; Reinhardt, A.; Landan, G.; Tria, F.D.K.; Turner, J.M.; Davies, C.; Schonheit, P. New views on an old enzyme: Allosteric regulation and evolution of archaeal pyruvate kinases. *FEBS J.* **2019**, *286*, 2471–2489. [CrossRef] [PubMed]

47. Pickl, A.; Johnsen, U.; Schönheit, P. Fructose degradation in the haloarchaeon *Haloferax volcanii* involves a bacterial type phosphoenolpyruvate-dependent phosphotransferase system, fructose-1-phosphate kinase, and class II fructose-1,6-bisphosphate aldolase. *J. Bacteriol.* **2012**, *194*, 3088–3097. [CrossRef] [PubMed]

48. Walden, H.; Taylor, G.; Lilie, H.; Knura, T.; Hensel, R. Triosephosphate isomerase of the hyperthermophile *Thermoproteus tenax*: Thermostability is not everything. *Biochem. Soc. Trans.* **2004**, *32*, 305. [CrossRef] [PubMed]

49. Sutter, J.M.; Tastensen, J.B.; Johnsen, U.; Soppa, J.; Schonheit, P. Key enzymes of the semiphosphorylative Entner-Doudoroff pathway in the haloarchaeon *Haloferax volcanii*: Characterization of glucose dehydrogenase, gluconate dehydratase, and 2-keto-3-deoxy-6-phosphogluconate aldolase. *J. Bacteriol.* **2016**, *198*, 2251–2262. [CrossRef] [PubMed]

50. Matsumura, H.; Izui, K.; Mizuguchi, K. A novel mechanism of allosteric regulation of archaeal phosphoenolpyruvate carboxylase: A combined approach to structure-based alignment and model assessment. *Protein Eng. Des. Sel.* **2006**, *19*, 409–419. [CrossRef] [PubMed]

51. Kawakami, R.; Satomura, T.; Sakuraba, H.; Ohshima, T. L-proline dehydrogenases in hyperthermophilic archaea: Distribution, function, structure, and application. *Appl. Microbiol. Biotechnol.* **2012**, *93*, 83–93. [CrossRef]

52. Singh, H.; Arentson, B.W.; Becker, D.F.; Tanner, J.J. Structures of the PutA peripheral membrane flavoenzyme reveal a dynamic substrate-channeling tunnel and the quinone-binding site. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3389–3394. [CrossRef]

53. Botou, M.; Lazou, P.; Papakostas, K.; Lambrinidis, G.; Evangelidis, T.; Mikros, E.; Frillingos, S. Insight on specificity of uracil permeases of the NAT/NCS2 family from analysis of the transporter encoded in the pyrimidine utilization operon of Escherichia coli. *Mol. Microbiol.* **2018**, *108*, 204–219. [CrossRef] [PubMed]

54. Rawls, K.S.; Martin, J.H.; Maupin-Furlow, J.A. Activity and transcriptional regulation of bacterial protein-like glycerol-3-phosphate dehydrogenase of the haloarchaea in *Haloferax volcanii*. *J. Bacteriol.* **2011**, *193*, 4469–4476. [CrossRef] [PubMed]

55. Sherwood, K.E.; Cano, D.J.; Maupin-Furlow, J.A. Glycerol-mediated repression of glucose metabolism and glycerol kinase as the sole route of glycerol catabolism in the haloarchaeon *Haloferax volcanii*. *J. Bacteriol.* **2009**, *191*, 4307–4315. [CrossRef] [PubMed]

56. Liu, H.; Orell, A.; Maes, D.; van Wolferen, M.; Lindas, A.C.; Bernander, R.; Albers, S.V.; Charlier, D.; Peeters, E. BarR, an Lrp-type transcription factor in *Sulfolobus acidocaldarius*, regulates an aminotransferase gene in a beta-alanine responsive manner. *Mol. Microbiol.* **2014**, *92*, 625–639. [CrossRef] [PubMed]

57. Liu, H.; Wang, K.; Lindås, A.-C.; Peeters, E. The genome-scale DNA-binding profile of BarR, a β-alanine responsive transcription factor in the archaeon *Sulfolobus acidocaldarius*. *BMC Genom.* **2016**, *17*, 569. [CrossRef] [PubMed]

58. Okamura, H.; Yokoyama, K.; Koike, H.; Yamada, M.; Shimowasa, A.; Kabasawa, M.; Kawashima, T.; Suzuki, M. A structural code for discriminating between transcription signals revealed by the feast/famine regulatory protein DM1 in complex with ligands. *Structure* **2007**, *15*, 1325–1338. [CrossRef] [PubMed]

59. Dyall-Smith, M. *The Halohandbook*. 2019. Available online: https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mds.pdf (accessed on 26 September 2019).

60. Darnell, C.L.; Tonner, P.D.; Gulli, J.G.; Schmidler, S.C.; Schmid, A.K. Systematic discovery of archaeal transcription factor functions in regulatory networks through quantitative phenotyping analysis. *MSystems* **2017**, *2*, e00032-17. [CrossRef] [PubMed]

61. Martin, J.H.; Sherwood Rawls, K.; Chan, J.C.; Hwang, S.; Martinez-Pastor, M.; McMillan, L.J.; Prunetti, L.; Schmid, A.K.; Maupin-Furlow, J.A. GlpR is a direct transcriptional repressor of fructose metabolic genes in *Haloferax volcanii*. *J. Bacteriol.* **2018**, *200*, e00244-18. [CrossRef]

62. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [CrossRef]

63. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.

64. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef]

65. Jiang, M.; Anderson, J.; Gillespie, J.; Mayne, M. uShuffle: A useful tool for shuffling biological sequences while preserving the k-let counts. *BMC Bioinform.* **2008**, *9*, 192. [CrossRef] [PubMed]

66. Fu, X.; Liu, R.; Sanchez, I.; Silva-Sanchez, C.; Hepowit, N.L.; Cao, S.; Chen, S.; Maupin-Furlow, J. Ubiquitin-like proteasome system represents a eukaryotic-like pathway for targeted proteolysis in archaea. *MBio* **2016**, *7*, e00379-16. [CrossRef] [PubMed]